```
                              GenCore version 5.1.4_p5_4578
                              Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March 14, 2003, 09:23:09 ; Search time 6.69057 Seconds
                 (without alignments)
                 2442.495 Million cell updates/sec

Title: .
Perfect score:   2120
Sequence:        1 MTPPPGRRAPSAPRARVLS..........TFQVADSHPEVAQRVEPMGP 394

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters:          112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

```
Result                        %      Query
No.     Score      Match    Length  DB   ID           Description
-----   -----      -----    ------  --   --           -----------
 1      2115       99.8      664    1    SM7A_MOUSE    Q9UR8 mus musculu
 2      1897       89.5      666    1    SM7A_HUMAN    075326 homo sapien
 3      389.5      18.4      861    1    SM4D_MOUSE    O09126 mus musculu
 4      368        17.4      749    1    SM3B_HUMAN    Q13214 homo sapien
 5      363        17.1      771    1    SM3A_HUMAN    Q14563 homo sapien
 6      361.5      17.1      748    1    SM3B_MOUSE    O62177 mus musculu
 7      361.5      17.1      862    1    SM4D_HUMAN    Q92854 homo sapien
 8      359        16.9      772    1    SM3A_MOUSE    O08665 mus musculu
 9      355.5      16.8      782    1    SM4B_MOUSE    O62179 mus musculu
10      354        16.7      772    1    SM3A_RAT      O63548 rattus norv
11      350        16.5      772    1    SM3A_CHICK    O09607 gallus gall
12      349        16.5      860    1    SZ1A_BRARE    Q9w7j1 brachydanio
13      344.5      16.2      766    1    SM27_BRARE    Q9yhx4 brachydanio
14      342.5      16.2      778    1    SZ1B_BRARE    Q9w686 brachydanio
15      341.5      16.1      441    1    VA39_VACCV    P24764 vaccinia vi
16      340.5      16.1      403    1    VA39_VACCC    P21062 vaccinia vi
17      340        16.0      775    1    SM3E_HUMAN    O15041 homo sapien
18      332.5      15.7      775    1    SM3E_MOUSE    P70275 mus musculu
19      331.5      15.6      751    1    SM3C_MOUSE    O62181 mus musculu
20      329.5      15.5      764    1    SM22_BRARE    Q9wg6 brachydanio
21      329.5      15.5      785    1    SM3E_CHICK    O42237 gallus gall
22      326.5      15.4      770    1    SM4F_HUMAN    O95754 homo sapien
23      324.5      15.3      751    1    SM3C_HUMAN    Q99985 homo sapien
24      324.5      15.3      777    1    SM4F_MOUSE    Q92123 mus musculu
25      318.5      15.0      751    1    SM3C_CHICK    O42236 gallus gall
26      316        14.9      673    1    SM4B_HUMAN    Q9npr2 homo sapien
27      314        14.8      834    1    SM4C_MOUSE    Q64151 mus musculu
28      313.5      14.8      761    1    SM3D_CHICK    O09663 gallus gall
29      313.5      14.8      838    1    SM4G_HUMAN    Q9ntn9 homo sapien
30      312        14.7      730    1    SM1A_SCHAM    O26473 schistocerc
31      309        14.6      776    1    SM4F_RAT      Q9143 rattus norv
32      307.5      14.5      785    1    SM3F_HUMAN    Q13275 homo sapien
33      303.5      14.3      777    1    SM3D_HUMAN    O95025 homo sapien
34      303        14.3      785    1    SM3F_MOUSE    O88632 mus musculu
35      303        14.3      888    1    SM6A_MOUSE    O35464 mus musculu
36      298.5      14.1      837    1    SM4G_MOUSE    Q9wuh7 mus musculu
37      297.5      14.0      712    1    SM1A_TRICF    Q26972 tribolium c
38      296.5      14.0      712    1    SM1A_CAEEL    O17330 caenorhabdi
39      292        13.8      697    1    SM2A_SCHGR    Q9xzc8 schistocerc
40      287.5      13.6      760    1    SM4A_MOUSE    O62178 mus musculu
41      275        13.0      1093   1    SM5B_MOUSE    O60519 mus musculu
42      274.5      12.9      1074   1    SM5B_HUMAN    Q13591 homo sapien
43      269.5      12.7      771    1    SM1A_DROME    Q24322 drosophila
44      261.5      12.3      762    1    SM4A_HUMAN    Q9h3s1 homo sapien
45      261        12.3      1077   1    SM5A_MOUSE    O62217 mus musculu
```

ALIGNMENTS

RESULT 1
SM7A_MOUSE

```
ID   SM7A_MOUSE     STANDARD;     PRT;     664 AA.
AC   Q9QUR8; O88371j
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 7A precursor (Semaphorin L) (Sema L) (Semaphorin K1) (Sema
DE   K1) (CD108 antigen) (CDw108).
GN   SEMA7A OR SEMK1 OR SEMAL OR CD108.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=9096477; PubMed=9878861;
RA   Sato Y., Takahashi H.;
RL   "Molecular cloning and expression of murine homologue of semaphorin K1
RL   gene.";
RL   Biochim. Biophys. Acta 1443:419-422(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=98389619; PubMed=9721204;
RA   Lange C., Liehr T., Goen M., Gebhart E., Fleckenstein B., Ensser A.;
RL   "New eukaryotic semaphorins with close homology to semaphorins of DNA
RL   viruses.";
RL   Genomics 51:340-350(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=20340072; PubMed=10885563;
RA   Mine T., Harada K., Matsumoto T., Yamana H., Shirouzu K., Itoh K.,
RA   Yamada A.;
RL   "CDw108 expression during T-cell development.";
RL   Tissue Antigens 55:429-436(2000).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE NERVOUS SYSTEM AND IN
CC       MODULATING IMMUNE FUNCTION.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, LOWER IN HEART,
CC       THYMUS, SPLEEN, TESTIS AND OVARY. THE EXPRESSION INCREASES IN LATE
CC       EMBRYONIC AND POSTNATAL STAGES.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; AB017532; BAA75665.1; -.
DR   EMBL; AF030699; AAC34262.1; -.
DR   EMBL; AF176670; AAD53118.1; -.
DR   MGD; MGI:13306826; Sema7a.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF01403; Sema; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein; GPI-anchor.
FT   SIGNAL        1     44       POTENTIAL.
FT   CHAIN        45    646       SEMAPHORIN 7A.
FT   PROPEP      647    664       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       75    474       SEMA.
FT   DOMAIN      557    618       IG-LIKE C2-TYPE DOMAIN.
FT   SITE        265    267       CELL ATTACHMENT SITE (POTENTIAL).
FT   LIPID       646    646       GPI-ANCHOR (POTENTIAL).
FT   DISULFID    564    611       BY SIMILARITY.
FT   CARBOHYD    102    102       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    154    154       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    600    600       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CONFLICT    357    357       P -> S (IN REF. 2).
SQ   SEQUENCE   664 AA;  74993 MW;  1804A31C3C55B715 CRC64;

Query Match             99.8%;
Best Local Similarity   99.7%;
Matches   393;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Score 2115;  DB 1;  Length 664;  Pred. No. 1.8e-172;

Qy    1  MTPPPGRRAAPSAPRARVLSLPAREFGLPLRLRLLLVFWVAAAASAQGHSRSGPRISAVWKG   60
         ||||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||
Db    1  MTPPPGRRAAPSAPRARVLSLPAREGLPLRLRLLIVFWVAAAASAQGHSRSGPRISAVWKG   60

Qy   61  QDHVDFSQPEPHTVLFHEPGSFSVWVGGRGKVYHFNFPEGKNASVRTVNIGSTKGSCQDK  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  QDHVDFSQPEPHTVLFHEPGSFSVWVGGRGKVYHFNFPEGKNASVRTVNIGSTKGSCQDK  120

Qy  121  QDCGNYITLLERRGNGLLVCGTNARKPSCWNLVNDSVVMSLGEMKGYAPFSPDENSLVLF  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  QDCGNYITLLERRGNGLLVCGTNARKPSCWNLVNDSVVMSLGEMKGYAPFSPDENSLVLF  180

Qy  181  EGDEVYSTIRKQEYNGKIPIRFRIRGESELYTSDTVMQNPQFIKATIVHQDQAYDDKIYY  240
         |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  181  EGDEVYSTIRKQEYNGKIPIRFRIRGESELYTSDIVMQNPQFIKATIVHQDQAYDDKIYY  240

Qy  241  FFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCSDAATNRNFNR  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  FFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCSDAATNRNFNR  300

Qy  301  LQDVFLLPDPSGQWRDTRVYGVFSNPWNYSAVCVYSLGDIDRVERTSSLKGYHMGLSNPR  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  301  LQDVFLLPDPSGQWRDTRVYGVFSNPWNYSAVCVYSLGDIDRVERTSSLKGYHMGLPNPR  360

Qy  361  PGMCLPKQPIPIETFQVADSHPEVAQRVEPMGP  394
         ||||||||||||||||||||||||||||||||
Db  361  PGMCLPKQPIPIETFQVADSHPEVAQRVEPMGP  394
```

```
RESULT 2
SM7A_HUMAN
ID   SM7A_HUMAN   STANDARD;   PRT;   666 AA.
AC   O75326;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 7A precursor (Semaphorin L) (Semaphorin
DE   K1) (Sema K1) (John-Milton-Hargen human blood group Ag) (JMH blood
DE   group antigen) (CD108 antigen) (CDw108).
GN   SEMA7A OR SEMAL OR CD108.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98389619; PubMed=9721204;
RA   Lange C., Liehr T., Goen M., Gebhart E., Fleckenstein B., Ensser A.;
RT   "New eukaryotic semaphorins with close homology to semaphorins of DNA
RT   viruses.";
RL   Genomics 51:340-350(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99218431; PubMed=10201933;
RA   Yamada K., Kubo K., Takeshita T., Harashima N., Kawano K., Mine T.,
RA   Sagawa K., Sugamura K., Itoh K.;
RT   "Molecular cloning of a glycosylphosphatidylinositol-anchored molecule
RT   CDw108.";
RL   J. Immunol. 162:4094-4100(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=98804063; PubMed=9712866;
RA   Xu X., Ng S., Wu Z.-L., Nguyen D., Homburger S., Seidel-Dugan C.,
RA   Ebens A., Luo Y.;
RT   "Human semaphorin K1 is glycosylphosphatidylinositol-linked and
RT   defines a new subfamily of viral-related semaphorins.";
RL   J. Biol. Chem. 273:22428-22434(1998).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=99344620; PubMed=10416131;
RA   Angelisova P., Drbal K., Cerny J., Hilgert I., Horejsi V.;
RT   "Characterization of the human leukocyte GPI-anchored glycoprotein
RT   CDw108 and its relation to other similar molecules.";
RL   Immunobiology 200:234-245(1999).
CC   -!- FUNCTION: MAY PLAY AN IMORTANT ROLE IN THE NERVOUS SYSTEM AND IN
CC       MODULATING IMMUNE FUNCTION.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, TESTIS, OVARY, SPLEEN,
CC       BRAIN, SPINAL CHORD, LUNG, HEART, ADRENAL GLAND, LYMPH NODES,
CC       THYMUS, INTESTINE AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------
DR   EMBL; AF030698; AAC34261.1; -.
DR   EMBL; AF030697; AAC34741.1; -.
DR   EMBL; AF069493; AAC82642.1; -.
DR   EMBL; AF071542; AAC80456.1; -.
DR   Genew; HGNC:10741; SEMA7A.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF01403; Sema; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein; GPI-anchor.
FT   SIGNAL        1     44       POTENTIAL.
FT   CHAIN        45    648       SEMAPHORIN 7A.
FT   PROPEP      649    666       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       75    474       SEMA.
FT   DOMAIN      559    620       IG-LIKE C2-TYPE DOMAIN.
FT   SITE        267    269       CELL ATTACHMENT SITE (POTENTIAL).
FT   LIPID       648    648       GPI-ANCHOR (POTENTIAL).
FT   DISULFID    566    613       BY SIMILARITY.
```

```
FT   CARBOHYD      105    105   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD      157    157   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD      258    258   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD      330    330   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD      602    602   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD      666    666   N-LINKED (GLCNAC...) (PROBABLE).
SQ   SEQUENCE      666 AA;  74823 MW;  AD3ABE56B5EBB194 CRC64;
```

```
Query Match             89.5%;  Score 1897;  DB 1;  Length 666;
Best Local Similarity   90.2%;  Pred. No. 6.9e-154;
Matches  358;  Conservative   8;  Mismatches  27;  Indels   4;  Gaps   2;

Qy     1 MTPPPGPGRAAPSAPRARVLSLPARFGPLPLRLRLLLVFWVAAASAQGHSRSGPRISAVWK- 59
           ||||||||||||||||||   ||  ||||||||||||||:||||||||||||: |||:|||
Db     1 MTPPPGPGRAAPSAPRARVPGPPARLGLPLRLRLLLWAAAASAQGHLRSGPRIFAVWKG  60

Qy    60 --GQDHVDFSQEPHTVLFHEPGSSFSVWWVGGRGKVYHFNFPEGKNASVRTVNIGSTKGSC 117
            ||||| ||: |||||||||||||| |:|||||||||:|:|||||||||||||||||||||
Db    61 HVGQDRVDFGQTEPHTVLFHEPGSSSVWWGGRGKVYLFDFPEGKNASVRTVNIGSTKGSC 120

Qy   118 QDKQDCGNYITLLERRGNGLLVCGTNARKPSCWNLVNDSVVMSLGEMKGYAPFSPDENSL 177
           :|: |:|||||||| ||.|:||:||||||:|||||||||:.::|.| .:||||||||||||
Db   121 LDKRDCENYITLLERRSEGLLACGTNARHPSCWNLVNGTVV-PLGEMRGYAPFSPDENSL 179

Qy   178 VLFEGDEVYSTIRKQEYNGKIPRFRRIRGESELYTSDTVMQNPQFIKATIVHQDQAYDDK 237
           |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db   180 VLFEGDEVYSTIRKQEYNGKIPRFRRIRGESELYTSDIVMQNPQFIKATIVHQDQAYDDK 239

Qy   238 IIYFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCSDAATNRN 297
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :|
Db   240 IIYFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCSDAATNKN 299

Qy   298 FNRLQDVFLLPDPSGQWRDTRVYGVGFSNPWNYSAVCVYSLGDIDRVFRTSSLKGYHMGLS 357
           ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||| .|
Db   300 FNRLQDVFLLPDPSGQWRDTRVYGVGFSNPWNYSAVCVYSLGDIDKVFRTSSLKGYHSSLP 359

Qy   358 NPRPGMCLPKKQPIPTETFQVADSHPEVAQRVEPMGP 394
           ||||| |||: |||||||||||| ||||||||||| ||
Db   360 NPRPGKCLPDQQPIPTETFQVADRHPEVAQRVEPMGP 396
```

```
DR   EMBL; U69535; AAC52964.1; -.
DR   MGD; MGI:109244; Sema4d.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF01403; Sema; 1.
DR   Pfam; PF01437; PSI; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00423; PSI; 1.
KW   Signal; Transmembrane; Immunoglobulin domain; Multigene family;
KW   Neurogenesis; Developmental protein; Glycoprotein.
FT   SIGNAL          1     23
FT   CHAIN          24    861   SEMAPHORIN 4D.
FT   DOMAIN         24    733   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      734    754   POTENTIAL.
FT   DOMAIN        755    861   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        228    523   SEMA.
FT   DOMAIN        569    630   IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID      576    624   BY SIMILARITY.
FT   CARBOHYD       49     49   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       77     77   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      139    139   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      191    191   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      379    379   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      419    419   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      613    613   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      632    632   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE      861 AA;  95714 MW;  533CD6D27JA6D79B CRC64;
```

```
Query Match             18.4%;  Score 389.5;  DB 1;  Length 861;
Best Local Similarity   29.3%;  Pred. No. 3e-25;
Matches  128;  Conservative  59;  Mismatches 151;  Indels  99;  Gaps  20;

Qy    24 RFGLPLR---LRLLLVFWVAAASAQGHSRSGPRISAVWKGQDHVDFSQEPHTVLFHEPG  80
           ||| |     |:|||   ||:   |:           ||:|:  |  |::|:|::| ||
Db     2 RMCAPVRGLFLALVVVLRTAVAFA----PVPRLT--WE----HGEVGLVQFHKPG  46

Qy    81 SF------SVWVGGRGKVYHFNFPEGKNASVRTVNIG------STKGS 116
           ::|     |:||||||||:|                      STKGS
Db    47 IFNYSALLMSEDKDTLYVGAREAVF----------AVNALNISEKQHEVAVKVSEDKKSK  96

Qy   117 CQDK----QDCGNYITLLE-RRGNGLLVCGTNARKPSC--WNLVNDSVVMSLGEMKGYA 168
           | ||    ::| |||| :|  |  :|:|||||| |  |            :||
Db    97 CAEKGRSKQTECLNYIRVLQPLSSTSLYVCGTNAFQPTCDHLNLTSFKFLGKSEDGKGRC 156

Qy   169 PFSPDENSLVLFEGDEVYSTIRKQEYN--GKIPRFRRIRGESELYTSDTV--MONPQFIK 224
           || |   : :::||:| |   |:  :   |        ::|:  :  :       |:|:|
Db   157 PFDPAHSYTSVMVGGELYS---GTSYNFLGSEPIISRNSSHSPLRTEYAIPWLNEPSFVF 213

Qy   225 ATIVHQD----QAYDKIYYFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWN 280
           |:|  .     ||:||:|:||  .|  :    : :  ||  |||||||| |  | ||||
Db   214 ADVIOKSPDGPEGEDDKVYFFTEVSVEYEFVFKLMIPRVARVCKGDQGGLRTLQ-KKWT 272

Qy   281 TFLKAMLVCSDAATNRNFNRLQDVFLLPDPSGQWRDTRVYGVFS---NPWNYSAVCVYSL 337
           :|||  | || :   :::  ||||| | | .  :  :  :     |        :||
Db   273 SFLKARLICSKPDSGLVFNILQDVFVLRAPG--LKEPVFYAVFTPQLNNVGLSAVCAYTL 330

Qy   338 GDIDRVF---------RTSSLK--GYHMGLSNPRPGMCLPKKQ--------PIPTET 375
           .:::::         |:||||  ||::  . :|||: ||                 PIPTET
Db   331 ATVEAVFSRGKYMQSATVEQSHTKWVRYNGPVPTRPGACIDSEARAANTSSLNLPDKT 390

Qy   376 FQVADSHPEVAQRVEPM 392
           | ||:   | |
Db ...
```

```
RESULT 3

ID   SM4D_MOUSE     STANDARD;      PRT;   861 AA.
AC   O09126;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 4D precursor (Semaphorin J) (Sema J) (Semaphorin C-like 2)
DE   (M-Sema G).
GN   SEMA4D OR SEMAJ OR SEMACL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=97125976; PubMed=8969198;
RA   Furuyama T., Inagaki S., Kosugi A., Noda S., Saitoh S.-I., Ogata M.,
RA   Iwahashi Y., Miyazaki N., Hamaoka T., Tohyama M.;
RT   "Identification of a novel transmembrane semaphorin expressed on
RT   lymphocytes.";
RL   J. Biol. Chem. 271:33376-33381(1996).
CC   -!- FUNCTION: MAY PLAY A FUNCTIONAL ROLE IN THE IMMUNE SYSTEM, AS WELL
CC       AS IN THE NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN LYMPHOID TISSUES,
CC       ESPECIALLY IN THE THYMUS, AS WELL AS IN THE NERVOUS TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
Db    391 LQFVKDHPLMDDSVTPI 407
```

RESULT 4
SM3B_HUMAN

```
ID   SM3B_HUMAN     STANDARD;      PRT;   749 AA.
AC   Q13214; Q93018;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 3B precursor (Semaphorin V) (Sema V).
GN   SEMA3B OR SEMA5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96210603; PubMed=8633026;
RA   Sekido Y., Bader S., Latif F., Chen J.-Y., Duh F.-M., Wei M.-H.,
RA   Albanesi J.P., Lee C.-C., Lerman M.I., Minna J.D.;
RT   "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
RT   cancer deletion region and demonstrate distinct expression patterns.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4120-4125(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dante M., Wamsley P.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
    SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS (BY
    SIMILARITY).
-!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY). ACCUMULATES IN THE
    ENDOPLASMIC RETICULUM.
-!- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY BUT DIFFERENTIALLY IN A
    VARIETY OF NEURAL AND NONNEURAL TISSUES.
-!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
-!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
-------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-------------------------------------------------------------------
DR   EMBL; U28369; AAD09118.1; -.
DR   EMBL; U73167; AAC02731.1; -.
DR   Genew; HGNC:10724; SEMA3B.
DR   MIM; 601281; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF01403; Sema; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Glycoprotein.
FT   SIGNAL        1     24
FT   CHAIN        25    749       SEMAPHORIN 3B.
FT   DOMAIN      239    537       SEMA.
FT   DOMAIN      637    717       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      699    702       POLY-GLY.
FT   DOMAIN      724    744       ARG-RICH (BASIC).
FT   DISULFID    644    710       BY SIMILARITY.
FT   CARBOHYD     82     82       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    124    124       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    427    427       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     29     36       PRLRLSPQ -> HAFGSPSKV (IN REF. 2).
SQ   SEQUENCE   749 AA;  83121 MW;  1F3BBF63F59A44F3 CRC64;
```

```
Query Match   17.4%;  Score 368;  DB 1;  Length 749;
Best Local Similarity   28.9%;  Pred. No. 1.7e-23;
Matches  127;  Conservative  49;  Mismatches  167;  Indels  96;  Gaps  18;

Qy    7 GRAAPSAPRARVLSLPARPGLPLRURLLLVFRWVAAASAQGHSRSGPRISAVWKGQDHVDF 66
          ||||||| |||  ||                                :  ||||
Db   22 GSAAPSPPPLRL-----------------------------------SPQELQA-WHGLQTFSL 49

Qy   67 SQPEPHTVLFHEPGSFSVWWGGRGKVYHFNFPEGGKNASVRTVNI----------------G 111
         ||   ::|||        ::|| |      |::|                            |
Db   50 ERTCYQALLVDEERGRLFVGAENIVASLNL---DNISKRAKKLAWPARVEWREECNWAG 106

Qy  112 STKGSCCQDKQDCGNYITLLER-RGNGLLVCGTNARKPSC-----WNLVNDSVVM----SL 161
         :|| ||    | |   :  || ||| |||||||  : ||         :|:  ::    ||
Db  107 KDIGT-----ECMNFVKLLHAYNRTHLLACGTGAFHPTCAFVEVGHRAEEPVLRLDPGRI 161

Qy  162 GEMKGYAPFSPDENSLVLFEGDEVYSTIRKQEYNGKIPRFRRIRGESELYT---SDTVMQ 218
          : ||:|| |||:|  :||:|||:|| :   : ||:||||||||||||||     |||:|
Db  162 EDGKGKSPYDPRHRAASVLVGEELYSGVAADLMGRDFTIFRSLGQRPSLRTEPHDSRWLN 221

Qy  219 NPQFIKA-TIVHQDQAYDDKIYFFREDNPDKNPE-APLNVSRVAQOLCRGDQGGESSLSV 276
          |:|:|  :|  ||||:||||||| |      ||  || |||| : :|| |  ||: || |
Db  222 EPKFVKVFWIPESENPDDDKIYFFFRETAVEAAAPALGRLSVSRVGQICRNDVGGQORSL-V 280

Qy  277 SKWNTFLKRAMLVCS--DAATNRNFNRLQDVFLLEPDPSGQWRDTRVYGVFSNP---WNYSA 331
          :||:|||| ::|||   || ||  | |||||||:    |   |||: |||||||     ::||
Db  281 NKWTTFLKARLVCSVPGVEGDTHFDLQDVFLL--SSRDHRTPLLIYAVFSTSSSSIFQGSA 338

Qy  332 VCVYSLGDIDRVF-------RTSSLKGYHMGLSNPRPGMCLPKK------QPIPTE 374
         |||||  :| |:|       | ||| || ||||||||||  :||       ||| |
Db  339 VCVYSMDVRRAFLGPFAHKEGPMHQWVSYQGRPYPRPGMC-PSKTFGTFSSSTKDFPDD 397

Qy  375 TFQVADSHPEVAQRVEPMG 393
          :|:|  :|: | : :|::|
Db  398 VIQFARNHPLMYNSVLPTG 416
```

RESULT 5
SM3A_HUMAN

```
ID   SM3A_HUMAN     STANDARD;      PRT;   771 AA.
AC   Q14563;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Semaphorin 3A precursor (Semaphorin III) (Sema III).
GN   SEMA3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=94094332; PubMed=8269517;
RA   Kolodkin A.L., Matthes D.J., Goodman C.S.;
RT   "The semaphorin genes encode a family of transmembrane and secreted
RT   growth cone guidance molecules.";
RL   Cell 75:1389-1399(1993).
RN   [2]
RP   SEQUENCE OF 1-37 FROM N.A.
RA   Woessner J., Minx P., Hinds K., Strowmatt C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 39-182 FROM N.A.
RA   Rohlfing T., Tin-Wollam A.M., Duckels G.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
    CONES. COULD SERVE AS A LIGAND THAT GUIDES SPECIFIC GROWTH CONES
    BY A MOTILITY-INHIBITING MECHANISM. BINDS TO THE COMPLEX
    NEUROPILIN-1/PLEXIN-1 (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Secreted (by similarity).
-!- DOMAIN: STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE CARBOXY
    THIRD OF THE PROTEIN.
```

```
CC  -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC  -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC  -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
DR  EMBL; L26081; AAA65938.1; -.
DR  EMBL; AC004451; AAC06185.1; -.
DR  EMBL; AC004848; AAC78622.1; -.
DR  MIM; 603961; -.
DR  Genew; HGNC:10723; SEMA3A.
DR  InterPro; IPR003599; Ig.
DR  InterPro; IPR003006; Ig_MHC.
DR  InterPro; IPR003659; Plexin-like.
DR  InterPro; IPR001627; Sema.
DR  Pfam; PF00047; ig; 1.
DR  Pfam; PF01403; Sema; 1.
DR  SMART; SM00409; IG; 1.
DR  SMART; SM00423; PSI; 1.
KW  Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW  Developmental protein; Glycoprotein.
FT  SIGNAL       1     20       POTENTIAL.
FT  CHAIN       21    771       SEMAPHORIN 3A.
FT  DOMAIN     240    538       SEMA.
FT  DOMAIN     642    729       IG-LIKE C2-TYPE DOMAIN.
FT  DOMAIN     727    769       ARG/LYS-RICH (BASIC).
FT  DISULFID   649    722       BY SIMILARITY.
FT  CARBOHYD    53     53       N-LINKED (GLCNAC...) (POTENTIAL).
FT  CARBOHYD   125    125       N-LINKED (GLCNAC...) (POTENTIAL).
FT  CARBOHYD   590    590       N-LINKED (GLCNAC...) (POTENTIAL).
SQ  SEQUENCE   771 AA;  88889 MW;  9985FBD3EAED08456 CRC64;

Query Match       17.1%;  Score 363;  DB 1;  Length 771;
Best Local Similarity  28.0%;  Pred. No. 4.8e-23;
Matches 119;  Conservative 66;  Mismatches 162;  Indels 78;  Gaps 19;

Qy    32  RLLLVFWVAAASAQGHSRSG-----PRISAVWK----GQDHVDF----SQPEHIVLFHEP   79
          :|| |:.:||            | |||:|||||    |||| :|     :|:|:|:|:||
Db     6  RIVCLFWGVLLTARANYQNGKNNVPRLKLSYKEMLESNNVIFPNGLANSSSYHFFLLDEE   65

Qy    80  GSFSVWVGGRGKVYHFNFPEGKNASVRTVNIGSTKGSCQDK--------QDCGNYITLL  130
          .: :|   :|.|| |:|:| :|||        |  :  |  :|       :||||:.:|
Db    66  RS-RLIYGAKDHIFSFDLVNIKDFQKIVWPVSYTR--RDECKWAGKDILKFCANFIKVL  121

Qy   131  ER-RGNGLLVCGTNARKPSC-----------WNLVNDSVVMSLGEMKGYAPFSPDENS  176
          :  .   :|.||| | | |             ||:.    :.|::.:| | | | .:.
Db   122  KAYNQTHLYACGTGAFHPICTYIEIGHHPEDNIFKLENS----HFENGRGKSPYDPKLIT  177

Qy   177  LVLFEGDEVYSTIRKQEYNGKIPRFRRIRGESELYT---SDTYMQNPQFIKATIVHQ-DQ  232
          :.|:.|:| |.  .  :.|::  |    : . :::    ||: :|:|:|| | :: | :
Db   178  ASLLIDGELYSGTAADFMGRDFAIFRTLGHHHPIRTEQHDSRWLNDPKFISAHLISESDN  237

Qy   233  AYDDKIYFFFREDNPDKNPEAPLNVSRVAQLCRGDQGESSLSVSKWNTFLKAMLVCSDA  292
          .:|||:||||||: .|: ..  ..  | .| | .| . :|| | ||::|||| |:|| :
Db   238  PEDDKVYFFFRENAIDGEHSGKATHARIGQICKNDFGGHRSL-VNKWTTFLKARLICSVP  296

Qy   293  ATN---RNFNRLQDVFLL--PDPSGQWRDTRVYGVF---SNPWNYSAVCVYSLGDIDRVF  344
          . |    : |::|||||:   ||    :|  ||. |   ||::: ||||:||:.|:|||
Db   297  GPNGIDTHFDELQDVFLMNFKDP----KNPVVGVFTTSSNIFKGSAVCMYSMSDVRVF  352

Qy   345  ------RTSSLKGYHMGLSNPRPGMCLPKK------QPIPTETFQVADSHPEVAQ  387
                | :: :||: :|.. | |.|   || :       :..  |:::|:|:|| :
Db   353  LGPYAHRDGPNYQWVPYQGRVPYPRGTC-PSKYFGGFEDSTKDLPDDVITFARSHPAMN  411

Qy   388  RVEPM  392
          .| ||
Db   412  PVFPM  416
```

```
RESULT 6
SM3B_MOUSE

ID  SM3B_MOUSE     STANDARD;      PRT;    748 AA.
AC  Q62177;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Semaphorin 3B precursor (Semaphorin V) (Sema A).
GN  SEMA3B OR SEMAA OR SEMA.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=NMRI; TISSUE=Embryo;
RX  MEDLINE=95267431; PubMed=7748561;
RA  Pueschel A.W., Adams R.H., Betz H.;
RT  "Murine Semaphorin D/collapsin is a member of a diverse gene family
RT  and creates domains inhibitory for axonal extension.";
RL  Neuron 14:941-948(1995).
CC  -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC      SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC  -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC  -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. LOW
CC      LEVELS FOUND BETWEEN DAYS 10-12. EXPRESSION PEAKS ON DAY 13 WITH
CC      MODERATE LEVELS FROM THEN UNTIL BIRTH.
CC  -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC  -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC  -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
DR  EMBL; X85990; CAA59982.1; -.
DR  MGD; MGI:1107561; Sema3b.
DR  InterPro; IPR003599; Ig.
DR  InterPro; IPR003006; Ig_MHC.
DR  InterPro; IPR003659; Plexin-like.
DR  InterPro; IPR001627; Sema.
DR  Pfam; PF00047; ig; 1.
DR  Pfam; PF01403; Sema; 1.
DR  SMART; SM00409; IG; 1.
DR  SMART; SM00423; PSI; 1.
KW  Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW  Developmental protein; Glycoprotein.
FT  SIGNAL       1     25       POTENTIAL.
FT  CHAIN       26    748       SEMAPHORIN 3B.
FT  DOMAIN     239    536       SEMA.
FT  DOMAIN     636    716       IG-LIKE C2-TYPE DOMAIN.
FT  DOMAIN     698    702       POLY-GLY.
FT  DOMAIN     723    743       ARG-RICH (BASIC).
FT  DISULFID   643    709       BY SIMILARITY.
FT  CARBOHYD    82     82       N-LINKED (GLCNAC...) (POTENTIAL).
FT  CARBOHYD   124    124       N-LINKED (GLCNAC...) (POTENTIAL).
SQ  SEQUENCE   748 AA;  82894 MW;  186GB7D2397C9305 CRC64;

Query Match       17.1%;  Score 361.5;  DB 1;  Length 748;
Best Local Similarity  29.9%;  Pred. No. 6.1e-23;
Matches 133;  Conservative 51;  Mismatches 152;  Indels 109;  Gaps 23;

Qy     7  GRAAPSAPRARV--LSLPARRGL-PLRL------RLLLV------FWVAAASAQ   45
          |  | .||| .:   |||  |  : ||||      |||||      |||||||||
Db    22  GDTAPNLPRLRSFQELQARHGVRFRLRERTCCYEALLVDEERGRLFVGAENHVASLSLD   81

Qy    46  GHSRSGPRISAVWKGQDHVDFSQPEHTVLFHEPGSFSVWVGGRGKVYHFNFPEGKNASV  105
```

```
Db   82  NISKRAKKLA--W-------------PAPVEWREECN---WAG----------  106
             :::  :                :  :
Qy  106  RTVNIGSTKGSCQDKDCGNYITLLERRGN-GLLVCGTNARKPSC----WNLVNDSVVMS  160
                      :  ::  :: ::: :: ::::::  : : : ::  :::::::::::
Db  107  --KDIGT--------ECMNFVRLLHAYNHTHLLACRTGAFHPTCALWRWATAGGTHAST  155

Qy  161  ----LGEMKGYTAPFSPDENSLVLFEGDEVYSTIRKQEYNGKIPRFRRIRGESELYT---S  213
             :  : :  : ::::::::::::::::: :  ::::::::::::::::::::  :  :
Db  156  GPEKLEDGKGKTPYDPRHRPPSVLVGEELYSGVTADLMGRDFTIFRSLGQNPSLRTEPHD  215

Qy  214  DTVMQNPQFIKA-TIVHDQQAYDDKIYYFREDNPDKNPE-APLNVSRVAQLCRGDOGGE  271
             :  ::::: :   :::::::  :: :::::::  :   : :::::::::::::::::::
Db  216  SRWLNEPKFVKVFWIPESENPDDDKIYFFRESAVEAAAPAMGRMSVSRVGQICRNDLGGQ  275

Qy  272  SSLSVSKWNTFLKAMLVCS--DAATNRNFNRLQDVFLLPDPSGQWRDT-RVYGVFSNP--  326
          :: :: ::::::: ::::    :::  :: :::::::::     ::::  ::: :: :
Db  276  RSL-VNKWTTFLKARLVCSVPGVEGDTHFDQLQDVFLL---SSRDRQTPLLYAVFSTSSG  331

Qy  327  -WNYSAVCVYSLGDIDRVF-----------RTSSLKGYHMGLSNPRPGMCLPKK------  368
            ::::::::::  : :::           ::::  :: :::::::::::::
Db  332  VFQGSAVCVYSMNDVRRAFLGPLPHKEGPTHQWVSYQGRVPVPRPGMC-PSKTFGTFSST  390

Qy  369  QPIPTETFQVADSHPEVAQRVEPMG  393
             : :   ::  :: ::  ::::::
Db  391  KDFPPDDVIQFGRNHPLMYNPVLPMG  415
```

RESULT 7

```
ID   SM4D_HUMAN     STANDARD;      PRT;   862 AA.
AC   Q92854;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 4D precursor (leukocyte activation antigen CD100) (BB18)
DE   (A8) (GR3).
GN   SEMA4D OR CD100.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=97030273; PubMed=8876214;
RA   Hall K.T., Boussiotis V.A., Dorfman D.M.,
RA   Bensussan A., Nadler L.M., Freeman G.J.;
RL   Proc. Natl. Acad. Sci. U.S.A. 93:11780-11785(1996).
RL   "Human CD100, a novel leukocyte semaphorin that promotes B-cell
RL   aggregation and differentiation.";
CC   -!- FUNCTION: MAY PLAY A FUNCTIONAL ROLE IN THE IMMUNE SYSTEM, AS WELL
CC       AS IN THE NERVOUS SYSTEM. INDUCES B CELLS TO AGGREGATE AND
CC       IMPROVES THEIR VIABILITY IN VITRO.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SKELETAL MUSCLE,
CC       PERIPHERAL BLOOD LYMPHOCYTES, SPLEEN, AND THYMUS AND ALSO
CC       EXPRESSED AT LOWER LEVELS IN TESTES, BRAIN, KIDNEY, SMALL
CC       INTESTINE, PROSTATE, HEART, PLACENTA, LUNG, AND PANCREAS BUT NOT
CC       IN COLON OR LIVER.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD100 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd100.htm".
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------
DR   EMBL; U60800; AAC50810.1; -.
DR   Genew; HGNC:10732; SEMA4D.
DR   MIM; 601866; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR001627; Sema.
KW   Signal; Transmembrane; Immunoglobulin domain; Multigene family;
KW   Neurogenesis; Developmental protein; Glycoprotein.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    862       SEMAPHORIN 4D.
FT   DOMAIN       22    734       SEMAPHORIN 4D.
FT   TRANSMEM    735    755       POTENTIAL.
FT   DOMAIN      756    862       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      228    523       SEMA.
FT   DOMAIN      569    630       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    576    624       BY SIMILARITY.
FT   CARBOHYD     49     49       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     77     77       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    191    191       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    379    379       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    632    632       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   862 AA;  96149 MW;  7B18EFEA98789371 CRC64;
```

```
Query Match             17.1%;          Score 361.5;   DB 1;   Length 862;
Best Local Similarity   27.7%;          Pred. No. 7.4e-23;
Matches  121;  Conservative   63;  Mismatches  154;  Indels   99;  Gaps  20;

Qy   24  RFGLPLR---LRLLLVFWVAAASAQGHSRSGPRISAVWKGQDHVDFSQPEPHTVLFHEPG   80
             ::    ::: : :     ::   :  :::  :                 :: ::::::
Db    2  RMCTPIRGLLMALAVMFGTAMAFA-----PIPRIT--WEHR------EVHLVQFHEPD    46

Qy   81  SF-------------SVWWGGRGKVYHFNFPEGGKNASVRTVNIG---------STKGS  116
         :               :  :  :  :  :                         ::::
Db   47  IYNYSALLLSEDKDTLYIGAREAVF--------AVNALNISEKQHEVYWKVSEDKKAK   96

Qy  117  CQDK-----QDCGNYITLLE-RRGNGLLVCGTNARKPSC--WNLVNDSVVMSLGEMKGYA  168
         :  :     : ::: :  :  :  : ::::::::::::    ::::::::::  : : :
Db   97  CAEKGKSKQTECLNYIRVLQPLSATSLYVCGTNAFQPACDHLNLTSFKFLGKNEDGKGRC  156

Qy  169  PFSPDENSLVLFEGDEVYSTIRKQEYN--GKIPRFRRIRGESELYTSDTV--MQNPQFIK  224
         ::    :: :: :: :: :    :     :  :   :   ::  :   :    ::  : :
Db  157  PFDPAHSYTSVWDGELYS---GTSYNFLGSEPIISRNSSHSPLRTEYAIPWLNEPSFVF   213

Qy  225  ATIVHQD----QAYDDKIYYFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWN   280
         :        : :: :  :: :    :     ::  :: :  :::::::  : :: ::
Db  214  ADVIRKSPDSPDGEDDRVYFFETEVSVEYEFVFRVLIPRIARVCKGDQGGLRTLQ-KKWT  272

Qy  281  TFLKAMLVCSDAATNRNFNRLQDVFLLPDPSGQWRDTRVYGVFS---NPWNYSAVCVYSL  337
         ::::: : ::  : ::   :: :::: : ::       :: :: :    :  ::::: : :
Db  273  SFLKARLICSRPDSGLVFNVLRDVFVLRSPG--LKVPVFYALFTPQLNNVGLSAVCAYNL  330

Qy  338  GDIDRVF----------RTSSLK--GYHMGLSNPRPGMCLPKKQ-----+-PIPTET   375
         :  ::            ::: :   :: ::::::::: :             :: ::
Db  331  STAEEVFSHGKYMQSTTVEQSHTKWVRYNGPVPKPRPGACIDSEARAANYTSSLNLPDKT  390

Qy  376  FQVADSHPEVAQRVEPM   392
         ::  :: ::  ::  :
Db  391  LQFVKDHPLMDDSVTPI   407
```

```
RESULT 8

ID   SM3A_MOUSE     STANDARD;      PRT;   772 AA.
AC   O08665; Q62180; Q62215;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 3A precursor (Semaphorin III) (Sema III) (Semaphorin D)
DE   (Sema D).
GN   SEMA3A OR SEMAD OR SEMD.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI; TISSUE=Embryo;
RX   MEDLINE=9267431; PubMed=7748561;
RA   Pueschel A.W., Adams R.H., Betz H.;
RT   "Murine semaphorin D/collapsin is a member of a diverse gene family
RT   and creates domains inhibitory for axonal extension.";
RL   Neuron 14:941-948(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9527432; PubMed=7748562;
RA   Messersmith E.K., Leonardo E.D., Shatz C.J., Tessier-Lavigne M.,
RA   Goodman C.S., Kolodkin A.L.;
RT   "Semaphorin III can function as a selective chemorepellent to pattern
RT   sensory projections in the spinal cord.";
RL   Neuron 14:949-959(1995).
RN   [3]
RP   SEQUENCE OF 107-772 FROM N.A.
RC   TISSUE=Fetal brain;
RA   Kimura T., Fishman M.C.;
RT   "cDNA sequence of mouse collapsin/semaphorin III.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9470885; PubMed=9331345;
RA   Taniguchi M., Yuasa S., Fujisawa H., Naruse I., Saga S., Mishina M.,
RA   Yagi T.;
RT   "Disruption of semaphorin III/D gene causes severe abnormality in
RT   peripheral nerve projection.";
RL   Neuron 19:519-530(1997).
CC   -!- FUNCTION: PLAY A ROLE IN GROWTH CONES GUIDANCE. MAY FUNCTION TO
CC       PATTERN SENSORY PROJECTIONS BY SELECTIVELY REPELLING AXONS THAT
CC       NORMALLY TERMINATE DORSALLY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN EMBRYONIC DEVELOPMENT
CC       (E11) IN DISTINCT REGIONS OF THE NEUROECTODERM AND MESODERM.
CC       EXPRESSION BECAME MORE EXTENSIVE AT LATER STAGES.
CC   -!- DOMAIN: STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE CARBOXY
CC       THIRD OF THE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
```

```
DR   EMBL; X85993; CAA59985.1; -.
DR   EMBL; D85028; BAA19773.1; -.
DR   EMBL; L41541; AAL77611.1; -.
DR   EMBL; L40484; AAA73934.1; -.
DR   MGD; MGI:107558; Sema3a.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF01403; Sema; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    772       SEMAPHORIN 3A.
FT   DOMAIN       40    538       SEMA.
FT   DOMAIN      240    730       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      643    728       ARG/LYS-RICH (BASIC).
FT   DOMAIN      728    770       ARG/LYS-RICH (BASIC).
FT   DISULFID     53    650       BY SIMILARITY.
FT   CARBOHYD    125    723       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     53     53       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    193    193       N-LINKED (GLCNAC...)
FT   CONFLICT    207    207       H -> D (IN REF. 4).
FT   CONFLICT    253    253       D -> N (IN REF. 1).
FT   CONFLICT    352    352       F -> S (IN REF. 4).
FT   CONFLICT    403    403       F -> L (IN REF. 1).
FT   CONFLICT    571    572       A -> G (IN REF. 1).
FT   CONFLICT    616    620       QH -> ED (IN REF. 1).
FT   CONFLICT    616    620       EDRKE -> RRSKR (IN REF. 1).
FT   CONFLICT    623    623       R -> K (IN REF. 4).
SQ   SEQUENCE   772 AA;  88799 MW;  E89A08528B1OAEC3 CRC64;

Query Match        16.9%;   Score 359;   DB 1;   Length 772;
Best Local Similarity   28.3%;   Pred. No. 1e-22;
Matches  119;  Conservative  62;  Mismatches  163;  Indels  76;  Gaps  19;

Qy    36  VFWVAAASAQGHSRSG----PRISAVWK----GQDHVDF----SQPEPHTVLFHEPGSFS   83
              |||                 : :||||    |:| || |       :| : |||:|||
Db    10  LFWGVLLTPARANYANGKNNVPRLKLSYKEMLESNNVITENGLANSSSYHFFLLDEERS-R   68

Qy    84  VWVGGRGKVYHFNPPEGKNASVRTVNIGSTKGSCQDK-----------QDCGNYITLLER-R  133
              |:||  ||       | |  :  |     |:  | |            :||| | |||: |
Db    69  LYVGAKDHIFSFNLVNIKDFQKIVWPVSYTR--RDECKWAGKDILKECANFIKVLEAYN   125

Qy   134  GNGLLVCGTNARKPSCWNL--------------VNDSVVMSLGEMKGYAPFSPDENSLVLFE  181
              |:|| |||:|  |:| |:|              |||| :|| | :|    | | |::|: |
Db   126  QTHLYACGTGAFHPICTYIEVGVHHPEDNIFKLQDS---HFENGRGKSPYDPKLLTASLLI   182

Qy   182  GDEYYSTIRKQEYNGKIPRFRRIRGESELYT---SDTVWMQNPQFIKATIV-HQDQAYDDK  237
              ::||:  ||:| |  |:| |:| ||:|| ||      :: | |:||:|    |:|:||  |
Db   183  DGEELYSGTRAADFMGRDFAIFRTLGHHHPIRTEQHDSRWLNDPRFISAHLIPESDNPEDDK  242

Qy   238  IYYFFREDNPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNNTFLKAMLVCSDAATN--  295
              ||  |::| |  || | |  ||  |||  |:|| || ||  ||||||:|| | ||| |
Db   243  VYFFFRENAIDGEHSGKATHARICQICKNDFGGHRSL-VNKWTTFLKARLICSVPGPNGI   301

Qy   296  -RNFNRLQDVFLL--PDPSGQWRDTRVGYGVF---SNPWNYSAVCYYSLGDIDRVE----   344
              : | :|||| :|: |: |   |   |: || |      | | |:| | || | ||
Db   302  DTHFDELQDVFLMNSKDP----KNPIVGVFTTSSNIFKGSAVCMYSMSDVRRVFLGPYA   357

Qy   345  ----RTSSLKGYHMGLSNPRPGMGLPKK---------QPIPTETFQVADSHPEVAQRVERM  392
              | :||  | | :| || ||:| ||           |:| | | || ||  |  |: |
Db   358  HRDGPNYQWVPYQGRVPYPRPGTC-PSKTFGGFDSTKDLPDDVITFARSHPAMYNPVFPI   416
```

```
RESULT 9

ID   SM4B_MOUSE     STANDARD;      PRT;   782 AA.
AC   Q62179;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Semaphorin 4B (Semaphorin C) (Sema C) (Fragment).
GN   SEMA4B OR SEMAC OR SEMC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI; TISSUE=Brain;
RX   MEDLINE=95267431; PubMed=7748561;
RA   Pueschel A.W., Adams R.H., Betz H.;
RT   "Murine semaphorin D/collapsin is a member of a diverse gene family
RT   and creates domains inhibitory for axonal extension.";
RL   Neuron 14:941-948(1995).
RN   [2]
RP   INTERACTION WITH GIPC.
RX   MEDLINE=99253973; PubMed=10318831;
RA   Wang L.-H., Kalb R.G., Strittmatter S.M.;
RT   "A PDZ protein regulates the distribution of the transmembrane
RT   semaphorin, M-SemF.";
RL   J. Biol. Chem. 274:14137-14146(1999).
CC   -!- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC       SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC   -!- SUBUNIT: INTERACTS WITH GIPC PDZ DOMAIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. LOW
CC       LEVELS FOUND BETWEEN DAYS 10-12. EXPRESSION PEAKS ON DAY 13 WITH
CC       MODERATE LEVELS FROM THEN UNTIL BIRTH.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
```

```
DR   EMBL; X85592; CAA59984.1; -.
DR   MGD; MGI:107559; Sema4b.
DR   InterPro; IPR003659; Sema4d.
DR   InterPro; IPR002165; Plexin-like.
DR   InterPro; IPR001627; Plexin_repeat.
DR   Pfam; PF01403; Sema; 1.
DR   Pfam; PF01437; PSI; 1.
DR   SMART; SM00423; PSI; 1.
KW   Transmembrane; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN       <1    662       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    663    683       POTENTIAL.
FT   DOMAIN      684    782       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      202    492       SEMA.
FT   DOMAIN      548    608       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      703    726       PRO-RICH.
FT   DISULFID    555    601       BY SIMILARITY.
FT   CARBOHYD     12     12       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     15     15       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     42     42       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     88     88       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    471    471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    526    526       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    574    574       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   782 AA;  86823 MW;  627A81FC8F8F7AC8 CRC64;

Qy   171 SPDENSLVLFEGDEVYSTIRKQEYNGKIPRFRIRGESELYTSDTV--MQNPOFIKAPIV 228
         :|| ||     | | | ||  |  ::  ||  :|:|| |  :  |    | || |
Db   134 DPNFKSTALVDGELY-TGTVSSFQGNDPAISRSQSSRPTKTESSLNWLQDPAFVASATS 192

Qy   229 HQDQAY-----DDKIYFFREDNPDKNPEAPLNVSRVAQLCRGDQOGESSLSVSKWNTFLK 284
         | |       ||||||| | |    | |:| |||||| ||  |   |   |  : |||||
Db   193 PESLGSPIGDDKIYFFFSETGQEFEFFENTIVSRVARVCKGDEGGERVLQ-QRWTSFLK 251

Qy   285 AMLVCSDAATNRFNFRLQDVFLL-PDPSGQWRDTRVVGVFSNPWNY-----SAVCVYSLG 338
         |  |||    | | :|||||| |  |  | |  | | |  ||      |   :| || |
Db   252 AQLLCSRPDDGPFPFNVLQDVFTLNPNPQ-DWRKTLSIGVFTSQWHRGTTEGSAICVFTMN 310

Qy   339 DIDRVF----------RTSSLKGYHMGLSNPRPGMCL 365
         | ::|           ||  |  | | ||  | | |
Db   311 DVQKAFDGLIYKKVNRETQQWYTETHQVPTRPGACI 346
```

Query Match   16.8%; Score 355.5; DB 1; Length 782;
Best Local Similarity   31.0%; Pred. No. 2.1e-22;
Matches 104; Conservative 47; Mismatches 138; Indels 47; Gaps 14;

```
Qy    72 HTVLFHEPGSFSVWWGGRGKVYHFN-----FPEGKNASVRTVNIGSTKGSC-----QDKQ 121
          :|| ||   :||| |        :|||:| ::||||   ||
Db    16 YTALLSQDGKTLVVGAREALFALNSNLSFLPGGEYQELLWSADADRKQQCSFKGKDPKR 75

Qy   122 DCGNYI-TLLERRGNGLLVCGTNARKPSC------WNLVND---SVVMSLGEMKGYAPF 170
          || ||| ||    |   | |          ||| ||  | :| ||  |
Db    76 DCQNYIKILLPLNSSHLLTCGTAAFSPLCAYIHIASFTLAQDEAGNVLLEDG--KGHCPF 133
```

RESULT 10
SM3A_RAT

```
ID   SM3A_RAT      STANDARD;      PRT;   772 AA.
AC   Q63548;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 3A precursor (Semaphorin III) (Sema III).
GN   SEMA3A.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=97073089; PubMed=8915837;
RA   Giger R.J., Wolfer D.P., De Wit G.M.J., Verhaagen J.;
RT   "Anatomy of rat semaphorin III/collapsin-1 mRNA expression and
RT   relationship to developing nerve tracts during neuroembryogenesis.";
RL   J. Comp. Neurol. 375:378-392(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN GUIDING GROWING AXONS TOWARDS THEIR
CC       TARGETS BY FORMING A MOLECULAR BOUNDARY THAT INSTRUCTS AXONS TO
CC       ENGAGE IN THE FORMATION OF SPECIFIC NERVE TRACTS. BINDS TO
CC       NEUROPILIN.
CC   -!- SUBCELLULAR LOCATION: Secreted (by similarity).
CC   -!- DEVELOPMENTAL STAGE: AT E-11, EXPRESSION WAS RESTRICTED TO THE
CC       OLFACTORY PIT, THE BASAL AND ROSTRAL SURFACE OF THE TELENCEPHALIC
CC       VESICLE, THE EYE ANLAGE, THE EPITHELIUM OF RATHKE'S POUCH, AND
CC       SOMITES. AT LATER DEVELOPMENTAL STAGES, IT WAS WIDELY DISTRIBUTED
CC       IN NEURONAL AS WELL AS IN MESENCHYMAL AND EPITHELIAL STRUCTURES
CC       OUTSIDE THE NERVOUS SYSTEM. AFTER BIRTH, MESENCHYMAL LEVELS
CC       DECREASED RAPIDLY AND EXPRESSION BECAME RESTRICTED TO SPECIFIC
CC       SETS OF NEURONS IN THE CNS. IN THE MATURE CNS, IT IS DETECTABLE IN
CC       MITRAL CELLS, NEURONS OF THE ACCESSORY OLFACTORY BULB AND CEREBRAL CORTEX,
CC       CEREBELLAR PURKINJE CELLS, AS WELL AS A SUBSET OF CRANIAL AND
CC       SPINAL MOTONEURONS.
CC   -!- DOMAIN: STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE CARBOXY
CC       THIRD OF THE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
```

```
DR   EMBL; X95286; CAA64607.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003659; Plexin-like.
```

Left column

```
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF01403; Sema; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    772       SEMAPHORIN 3A.
FT   DOMAIN      240    538       SEMA.
FT   DOMAIN      643    730       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      728    770       ARG/LYS-RICH (BASIC).
FT   DISULFID    650    723       BY SIMILARITY.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   772 AA;  88808 MW;  24090781 2FF9F2D2 CRC64;

Query Match             16.7%;  Score 354;  DB 1;  Length 772;
Best Local Similarity   28.1%;  Pred. No. 2.8e-22;
Matches  118;  Conservative  63;  Mismatches 163;  Indels  76;  Gaps  19;

Qy   36 VFWVAAASAQGHSRSG-----GQDHVDF----SQPEPHTVLFHEPGSFS 83
         :||            ||||           :| ||||
Db   10 LFWGILILTARANYANGKNNVPRLKLSYKEMLESNNVITFNGLANSSSYHTFLLDEERS-R 68

Qy   84 VWVGGRGKVYHFNFPEGKNASVRTVNIGSTKGSCQDK-------QDCGNYITLLER-R 133
Db   69 LYVGAKDHIFSFNLVNIKDFQKIVWPVSYTR---RDECKWAGKDILKECANFIKVLKAYN 125

Qy  134 GNGLLVCGTNARKPSCWNL-------VNDSVVMSLGEMKGYAPFSPDENSLVLFE 181
Db  126 QTHLYACGTGAFHPICTYIEVGHHPEDNIFKLQDS--HFENGRGKSPYDPKLLTASLLI 182

Qy  182 GDEVYSTIRKQEYNGKIPRFRRIRGESELYT----SDTVMQNPQFIKATIV-HQDQAYDDK 237
Db  183 DGELYSGTAADFMGRDFAIFRTLGHHHPIRTEQHDSRWLNDPRFISAHLIPESDNPEDDK 242

Qy  238 IIYFFREDNDPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCSDAATN-- 295
Db  243 VYFFFRENAIDGEHSGKATHARIGQICKNDFGGHRSL-VNKWTTFLKARLICSVPGPNGI 301

Qy  296 -RNFNRLQDVFLL--PDPSGQWRDTRVYGVF----SNPWNYSAVCVYSLGDIDRVF---- 344
Db  302 DTHFDELQDVFLMNSKDP---KNPIVYGVFTTSSNIFKGSAVCMYSMSDVRRVFLGPYA 357

Qy  345 ---RISSLKGYHMGLSNPRPGMCLPKK-----QPIPIETFQVADSHPEVAQRVEPM 392
Db  358 HRDGPNYQWVPYQGRVPYPRPGTC-PSKTFGGFDSTRKDLPDDVITFARSHPAMYNPVFPI 416
```

```
RESULT 11
SM3A_CHICK
ID   SM3A_CHICK     STANDARD;      PRT;   772 AA.
AC   Q90607;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 3A precursor (Collapsin-1) (COLL-1).
GN   SEMA3A OR COLL1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 322-329; 362-372; 395-403 AND
RP   666-680.
RC   TISSUE=Brain;
RX   MEDLINE=94065554; PubMed=8402908;
RA   Luo Y., Raible D., Raper J.A.;
RT   "Collapsin: a protein in brain that induces the collapse and paralysis
RT   of neuronal growth cones.";
```

Right column

```
RL   Cell 75,217-227(1993).
CC   -!- FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
CC       CONES. COULD SERVE AS A LIGAND THAT GUIDES SPECIFIC GROWTH CONES
CC       BY A MOTILITY-INHIBITING MECHANISM. BINDS TO NEUROPILIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT RELATIVELY HIGH LEVELS IN BRAIN
CC       AND MUSCLE, MODERATE LEVELS IN LUNG, BURSA, AND HEART AND
CC       VIRTUALLY ABSENT IN LIVER. COLLAPSIN-1, -2, -3, AND -5 BIND TO
CC       OVERLAPPING BUT DISTINCT AXON TRACTS.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- DOMAIN: STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE CARBOXY
CC       THIRD OF THE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U02528; AAC59638.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF01403; Sema; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    772       SEMAPHORIN 3A.
FT   DOMAIN      240    538       SEMA.
FT   DOMAIN      643    730       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      728    770       ARG/LYS-RICH (BASIC).
FT   DISULFID    650    723       BY SIMILARITY.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   772 AA;  88867 MW;  E91E09DE0CC940AC CRC64;

Query Match             16.5%;  Score 350;  DB 1;  Length 772;
Best Local Similarity   28.6%;  Pred. No. 6.1e-22;
Matches  123;  Conservative  56;  Mismatches 165;  Indels  86;  Gaps  19;

Qy   29 LRLRLLLVFWVAAASAQGHSRSGPRISAVWK----GQDHVDF----SQPEPHTVLFHEPG 66
Db    7 IALLSLGVLLAGRVNCQHVKKNNVPRLKLSYKEMLESNNVINFNGLANSSSYHTFLLDEER 66

Qy   81 SFSWVWGGRGKVYHFNFPEGKNASVRTVNL--------GSTKGSCQ-----DKQDCG 124
Db   67 S-RLYVGAKDHIFSFNL------VNIKEYQKIWPVSHSRRDECKWAGKDILRECA 115

Qy  125 NYITLLER-RGNGLLVCGTNARKPSCWNL----------VNDSVVMSLGEMKGYAPFS 171
Db  116 NFIKVLKTYNQTHLYACGTGAFHPMCTYIEVGSHPEDNIFRMEDS---HFENGRGKSPYD 172

Qy  172 PDENSLVLFEGDEVYSTIRKQEYNGKIPRFRRIRGESELYT---SDTVMQNPQFIKATIV 228
Db  173 PKLLTASLLVDGELYSGTAADFMGRDFAIFRTLGHHHPIRTEQHDSRWLNDPRFISAHLI 232

Qy  229 -HQDQAYDDKIYFFREDNDPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAML 287
Db  233 PESDNPEDDKIYFFFRENAIDGEHTGKATHARIGQICKNDFGGHRSL-VNKWTTFLKARL 291

Qy  288 VCSDAATN---RNFNRLQDVFLL--PDPSGQWRDTRVYGVF----SNPWNYSAVCVYSLGD 339
Db  292 ICSVPGPNGIDTHFDELQDVFLMNSKDP----KNPIVYGVFTTSSNIFKGSAVCMYSMSMTD 347
```

```
Qy   340  IDRVF-----RTSSLKGYHMGLSNPRPGMCLPKK-------QPIPTETFQVADSH  382
          |||  ||||| |||   | | |||| |    ::|||:||| ||
Db   348  VRRVFLGPYAHRDGPNYQWVPYQGRVPYPRPGTC-PSKTFGGFDSTKDLPDEVITFARSH  406

Qy   383  PEVAQRVEPM  392
          || ||:
Db   407  PAMNPVFPI  416
```

RESULT 12

```
ID   SZ1A_BRARE     STANDARD;      PRT;   860 AA.
AC   Q9W7J1;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Semaphorin Z1A precursor (Semaphorin 1A) (Sema-Z1A).
GN   SEMAZ1A OR SEMA3AA
OS   Brachydanio rerio (zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=99313409; PubMed=10386838;
RA   Yee C.S., Chandrasekhar A., Halloran M.C., Shoji W., Warren J.T.,
RA   Kuwada J.Y.;
RT   "Molecular cloning, expression, and activity of zebrafish semaphorin
RT   Z1a.";
RL   Brain Res. Bull. 48:581-593(1999).
CC   -!- FUNCTION: MAY INFLUENCE OUTGROWTH BY A VARIETY OF GROWTH CONES
CC       INCLUDING THOSE OF THE POSTERIOR LATERAL LINE GANGLION.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN HIGHLY SPECIFIC PATTERNS WITHIN
CC       THE DEVELOPING EMBRYO.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF086761; AAD43964.1; -.
DR   ZFIN; ZDB-GENE-991209-3; sema3aa.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR001627; Sema.
     Pfam; PF00047; ig; 1.
     Pfam; PF01403; Sema; 1.
     SMART; SM00409; IG; 1.
     SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    860       SEMAPHORIN Z1A.
FT   DOMAIN      241    539       SEMA.
FT   DOMAIN      645    724       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      722    858       ARG/LYS-RICH (BASIC).
FT   DISULFID    652    717       BY SIMILARITY.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    126    126       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    593    593       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   860 AA;  97263 MW;  5FD4C12194F5165C CRC64;
```

Query Match 16.5%; Score 349; DB 1; Length 860;

Best Local Similarity 26.6%; Pred. No. 8 6e-22;
Matches 122; Conservative 55; Mismatches 155; Indels 126; Gaps 21;

```
Qy     6  PGRAAPSAPRARVLSLPARFGLPLRLRLLLVFWVAAASAQGHS--RSGPRISAVWKGQD   62
          |||| |:   :|  | | ||  || |  | :|: ||| : |:: | :  |::::|  ||
Db    17  PGRVAPQHTKENVPRLKLSYNEMLESSNLVTF-TGLANSSGYDTFLMDGERGRLLVGAED   75

Qy    63  HVDFSQPEPHTVLFHEPGSFSVWVGGRGKVYHFNFPPGKNASVRTVNIG-------    111
          ||   :                              |:|               :
Db    76  HV-----FSFD------------------------------LVNINRDVKQIAWPAT    97

Qy   112  -STKGSCQ------DKQDCGNYITLER-RGNGLLVCGTNARKPSCWNLVNDSVVMSLGEM  164
           :|:  ||       |:| ||| | ::  || |||||||: ||| ||:|||::::|:| |
Db    98  PSKRDECKWAGRDLRKDCSNFVRVLQSYNQOTHIYICGTGAFHPIC------SFLEMGKR  150

Qy   165  ----KGYAPFSPDENSLVLFEGDEVYSTIRKQEYNGKIPRFRRIRGES              208
              :| |   |||:|||:|||||:|||||||||::|||||:|||:|||::
Db   151  AEDNIFRLDANYFENGRGKSPYDPKMQGSSLLLDGELYSGTISADFMGRDFAIFRTLGSHH  210

Qy   209  ELYT---SDTVMQNPQFIKATIV-HQDQAYDDKIYFFREDNPDKNPEAPLNVSRVAQLC   264
          ::|:   :| || |||:||| || ||:||:|||||||::::|:: |||:|||||:||||:
Db   211  PIRTEQHDSRWINEPRFGLIHIJIPESDNPEDDKIFLFFKENAMDGEHTGKAFISRIGQLC  270

Qy   265  RGDQGGESSLSVSKWNTFLKAMLVCSDAATN---RNFNRLQDVFLL--PDPSGQWRDTRV  319
           :| ||  ||  ||| |||||::||| :  |:   ||:||||||||:   |: |:  :|||
Db   271  KNDMGGHRSL-VNKWTFLKAKLTCSVPGLNGIDTHPDELQDVFLMSAKDP---KNPVI    325

Qy   320  YGVF---SNPWNYSAVCVYSLGDIDRVFRTSSLKGY-HMGLSN------------PRPGM  363
          |.||   |: ||||| | || |||:|||| |||||| |||||:            |||| |
Db   326  YAVFTTSSNIFRGSAICMYSMADIRRVF---LGPYAHRDGPNYQWVPFQGRVPYPRPGT   381

Qy   364  CLPKK------QPIPTETFQVADSHPEVAQRVEPMG  393
          |  |:       |||||||:||||  |:|: |:|||
Db   382  C-PSKTFGGFDSTKDLPDDVITFARLHPAMNPVQPMG  418
```

RESULT 13

```
ID   SM27_BRARE     STANDARD;      PRT;   766 AA.
AC   Q9YHX4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Semaphorin Z7 precursor (Semaphorin 7) (Sema-Z7).
GN   SEMAZ7 OR SEMA7
OS   Brachydanio rerio (zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99060909; PubMed=9867349;
RA   Halloran M.C., Severance S.M., Yee C.S., Gemza D.L., Kuwada J.Y.;
RT   "Molecular cloning and expression of two novel zebrafish
RT   semaphorins.";
RL   Mech. Dev. 76:165-168(1998).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF073289; AAC72345.1; -.
DR   ZFIN; ZDB-GENE-990715-7; sema7.
DR   InterPro; IPR003006; Ig_MHC.
```

```
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR001627; Sema.
     Pfam; PF00047; ig; 1.
     Pfam; PF01403; Sema; 1.
     Pfam; PF01437; PSI; 1.
     SMART; SM00410; IG_like; 1.
     SMART; SM00423; PSI; 1.
KW   Signal; Transmembrane; Immunoglobulin domain; Multigene family;
KW   Glycoprotein.
FT   SIGNAL        1     24
FT   CHAIN        25    766       POTENTIAL.
FT   DOMAIN       25    664       SEMAPHORIN 27.
FT   TRANSMEM    665    685       POTENTIAL.
FT   TRANSMEM    665    685       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      686    766       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      232    522       SEMA.
FT   DOMAIN      570    630       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    577    623       BY SIMILARITY.
FT   DISULFID     52     52
FT   CARBOHYD     52     52       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    433    433       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    766 AA;  85617 MW;  0ACBC693FE7DB30C CRC64;

Query Match          16.2%;   Score 344.5;   DB 1;   Length 766;
Best Local Similarity 29.8%;  Pred. No. 1.8e-21;
Matches  114;  Conservative  47;  Mismatches 146;  Indels  75;  Gaps 18;

Qy   75  LFHEPGSFSVW------------------VGGRGKVYHFNFPE--GKNASVRTVNIGSTKGSC 117
         ||||                         :||  |  | ||| |   | |:  || :|||||
Db   44  LFREEG---IWNTTMLLRDLNLILGAREAIFALDLDDITIKKAMLKWEVTRDQQNDC 100

Qy  118  QDK------QDCGNYITLLERRGNG-LIVCGTNARKPSCWNLVNDSVVMSL---GEMKGY 167
         | |      || ||||| ||| | | |||||||||||||||||| |   || ||
Db  101  SNGKDATNDCKNYIRILHKKNDGRMYVCGTKAFNPTCGYLSYADGKLTLEILQEDTKGK 160

Qy  168  APFSPDE---NSLVLFEGDEVYSTIRKQEYNGKIPRFRRIRGES--ELYTSDIYMQNPQF 222
         | | |      :|| ||| |  |  ||    |   :  |  ||  |:|  :|  |
Db  161  CPFDPFQRYTSANV----DGAYTSATSMNFRGSEPVVMRSTEESIRTEFTS-TWLSEPNF 215

Qy  223  IKATIVHQDQAY---------DDKIYYFFREDNDKNPEAPLNVSRVAQLCRGDQGGESSLSVSK 278
         |   ||||:          :||  |:|||:|||||||||  |:||| |  | |:  |||
Db  216  IHMAHIPEGQOSNPDGDDDKIYLFFSETAVEYEYSTKVDVSRVARVCKGDLGGQRTLQ-KK 274

Qy  279  WNFFLKAMLVCSDAATNRNFNRLQDVF-LLPDPSGQWRDTRVGVF----SNPWNYSAVCV 334
         || |||  || |  |  |   |  || | | :     |||    :|||:|:||| |
Db  275  WTSFLKARLDCQVPNTNLPL-LVQDVFHLCPD---DWTTCVFYAVFTPQSDSSQYSAVCS 330

Qy  335  YSLGDIDRVFRTSSLKG----------------YIMGLSNPRPGMC-----------LPKKQPIP 372
         |:| |  |   :| |                   | |  || ||||             |  | |:
Db  331  YKIEDIKIYFSKGFKAPFNVETSFVKWVMYSGELPDPRPGACIDNHAREKGITKSLELP 390

Qy  373  TETFQVADSHPEVAQRVEPMGP 394
         | :|      || |:  | |
Db  391  DKTLQFVDKPLMDQAVFAEQP 412
```

```
RESULT 14
ID   SZ1B_BRARE      STANDARD;   PRT;   778 AA.
AC   Q9W686;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Semaphorin 21B precursor (Semaphorin 1B) (Sema-21B).
GN   SEMA21B OR SEMA3AB.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99425174; PubMed=10495275;
RA   Roos M., Schachner M., Bernhardt R.R.;
RT   "Zebrafish semaphorin Z1b inhibits growing motor axons in vivo.";
RL   Mech. Dev. 87:103-117(1999).
CC   -!- FUNCTION: MIGHT NORMALLY INFLUENCE THE MIDSEGMENTAL PATHWAY CHOICE
CC       OF THE VENTRALLY EXTENDING MOTOR AXONS BY CONTRIBUTING TO A
CC       REPULSIVE DOMAIN IN THE POSTERIOR SOMITE.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN RHOMBOMERES THREE AND FIVE, AND
CC       IN THE POSTERIOR HALF OF NEWLY FORMED SOMITES WHICH IS AVOIDED BY
CC       VENTRALLY EXTENDING MOTOR AXONS.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF083382; AAD28103.1; -.
DR   ZFIN; ZDB-GENE-991209-6; sema3ab.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR001627; Sema.
     Pfam; PF00047; ig; 1.
     Pfam; PF01403; Sema; 1.
     Pfam; PF01437; PSI; 1.
     SMART; SM00409; IG; 1.
     SMART; SM00423; PSI; 1.
KW   Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein.
FT   SIGNAL        1     17
FT   CHAIN        18    778       POTENTIAL.
FT   DOMAIN       18    778       SEMAPHORIN 21B.
FT   DOMAIN      241    539       SEMA.
FT   DOMAIN      645    723       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      721    776       ARG/LYS-RICH (BASIC).
FT   DISULFID    652    716       BY SIMILARITY.
FT   CARBOHYD     54     54       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    127    127       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    593    593       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    778 AA;  88904 MW;  4D36F43234E21895 CRC64;

Query Match          16.2%;   Score 342.5;   DB 1;   Length 778;
Best Local Similarity 28.8%;  Pred. No. 2.7e-21;
Matches  124;  Conservative  64;  Mismatches 152;  Indels  91;  Gaps 25;

Qy   33  LLLVFWAAASAQG------HSRSG--PRISAVWK-----GODHVDF----SQPEHTVLFHEP 79
         ||| |             |:|  | :|| |          || ||     |  |  | |
Db    7  IVLLIWTLIAPERGTVAQRSKSNVPRLKPSYKEMLESNNLITPNGLANSSAYHTFLDEE 66

Qy   80  GSFSVWVGGRGKVYHFNFPEGKNASVRTVNIG----------STKGSCQ--:-DKODCG 124
         |||  |:| |||||||||||| | | | ||             |||:| |      ||:|
Db   67  RG-RLFVGAKIDVLSENLVD-------INMDQQLISWPSSPSRRDECKWAGKDVOKECA 117

Qy  125  NYITLLER-RGNGLLVCGTNARKPSCWNLV-------NDSVVMSLGEM----KGYAPFSPD-E 174
         ||| |:|  ||| | ||||||||||| |:       || | |:| |       |:| |:  |
Db  118  NFIKVLQPFNQTHLYACGTGAFHPVCAHVEVGKRSEDNTFRLGSSFENGRGKSPYDPKLQ 177

Qy  175  NSLVLFEGDEVYSTIRKQEYNGKIPRFRRIRGESELYT----SDTVMQNPQFIKATIV-HQ 230
         ||:|| ||| |  |  |:| | |   :: |  |:  |     |   |:| ||| | |  |
Db  178  TASMLIDG-ELYAGTSADFMGRDFAIFRTLGKHHPIRTEQHDSRWLNDPRVSVHLIPES 236

Qy  231  DQAYDDKIYFFREDNDPDKNPEAPLNVSRVAQLCRGDQGGESSLSVSKWNTFLKAMLVCS 290
         | | ||:|||| |:||||||||||||||||||||| | |  |:| |:   |  ||:|||
Db  237  DNAEDDKIYLFFRENAIDGEQISKAHARIGQLCKNDFGGHRSL-VNKWTTFLKARLVCS 295
```

```
Qy    291 DAATN---RNFNRLQDVFLL--PDPSGQWRDTRVYGVP---SNPWNYSAVCVYSLGDIDR 342
              ::|:|   |||||||:|  |||:|:||||||:||     ||||:||:|||||||:| ||
Db    296 VPGLNGIDTHPDELQDVFLMSSKDP----KNPIIYAVFTTSSNIFKGSSAVCMYSMADIRR 351

Qy    343 VFRTSSLKGY-HMGLSN------------PRPGMCLPKK------QPIPTETFQVADS 381
              ||:|||:|| | ||:||            ||||||||||      |:| |||:||:
Db    352 VF----LGPYAHRDGPNYQWVPFLNRVPYPRPCTC-PSKTFDGFPESTKDFPDDVITFARS 406

Qy    382 HPEVAQRVEPM 392
              ||:||||||||
Db    407 HPAMYNPVPPI 417
```

```
RESULT 15

ID     VA39_VACCV    STANDARD;    PRT;   441 AA.
AC     P24764;
DT     01-MAR-1992 (Rel. 21, Created)
DT     01-MAR-1992 (Rel. 21, Last sequence update)
DT     16-OCT-2001 (Rel. 40, Last annotation update)
DE     Protein A39.
GN     A39R OR SALI9R/SALF1R.
OS     Vaccinia virus (strain WR).
OC     Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC     Orthopoxvirus.
OX     NCBI_TaxID=10254;
RN     [1]
RP     SEQUENCE FROM N.A.
RX     MEDLINE=91259063; PubMed=2045793;
RA     Smith G.L., Chan Y.S., Howard S.T.;
RT     "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT     the right inverted terminal repeat.";
RL     J. Gen. Virol. 72:1349-1376(1991).
RN     [2]
RP     SEQUENCE FROM N.A.
RX     MEDLINE=91310644; PubMed=1856205;
RA     Amegadzie B.Y., Ahn B.-Y., Moss B.;
RT     "Identification, sequence, and expression of the gene encoding a Mr
RT     35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase.";
RL     J. Biol. Chem. 266:13712-13718(1991).
CC     -!- CAUTION: IN REF.2 THIS SEQUENCE IS IN TWO PART (SALI9R AND
CC         SALF1R) DUE TO THE PRESENCE OF FRAMESHIFT COMPARED TO THAT OF
CC         REF.1.
CC     --------------------------------------------------------------------------
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC     the European Bioinformatics Institute. There are no restrictions on its
CC     use by non-profit institutions as long as its content is in no way
CC     modified and this statement is not removed. Usage by and for commercial
CC     entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC     or send an email to license@isb-sib.ch).
CC     --------------------------------------------------------------------------
DR     EMBL; M61187; AAA48335.1; -.
DR     EMBL; D11079; BAA01811.1; ALT_SEQ.
DR     EMBL; X57318; CAA40587.1; -.
DR     PIR; S29921; S29921.
DR     InterPro; IPR001627; Sema.
DR     Pfam; PF01403; Sema; 1.
SQ     SEQUENCE   441 AA;  50185 MW;  F6FF8B6A6E6C9E50 CRC64;
```

```
Query Match             16.1%;   Score 341.5; DB 1; Length 441;
Best Local Similarity   31.1%;   Pred. No. 1.5e-21;
Matches     98; Conservative   51; Mismatches  127; Indels   39; Gaps  14;
```

```
Qy     84 VWVGGRGKVHFNFPEGKNASVRTVNIGSTKGSCQDKQDCGNYITL---LERRGNLLVC 140
             ::|:||   |:||:|:|     ||| || :||  || |||  |   ||| |||||||
Db     74 LYTGVNGAVTFS---NNKLNKT--GLTN------NNYITTSIKVEDADKDTLVC 117

Qy    141 GTNARKPSCWNLVNDSVVMSLGEMKGYAPFSPDENSLVLFEGDEVYSTIRKQEYNGKIPR 200
             :: ||||| ||| || ::  ::||    ||    ||::   :   || ||::  |||||
Db    118 GTNNGNPKCWKIDGSDPDKHRG---RGYAPYQNSKYTIISHNGC-VLSDINISKEG--IKR 172

Qy    201 FRRIRGES--ELYTSDTVMQNPQFIKATIVHQDQAYDDKIYYFFREDNPDKNPEAPLNVS 258
```

```
Db    173 WRRFEDPGPCGYDYLYTADNVIPK-DGLRGAFVDKDGTY-DKVYILFFDTIGSKR---IVKIP 227
             ::|:||:|  ||  || :: ||:  |:|: |  |: || | |:||:|:|| ||   | |:
Qy    259 RVAQLCRGDQGGESSLSVSKWNTELKAMLVCS-DAATNRNFNRLQDVFLLPDPSGQWRDT 317
             ::|  ||:||  |||| |:  |:|    |:|   |  |:||||:  |  ||| |||||||
Db    228 YIAQMCLNDEGGPSSLSSHRWSTFLKVELECDIDGGRSYRQI--IHSRTIKTD-----NDT 280

Qy    318 RVYGVFSNPWNYSAVCVYSLGDIDRVFRTSSLKGYVHMGLSNPRPGMCLPKKQPIPTETFQ 377
             |:|| : ||||:|||||||||:|  | :|| |:|  ||   | |||| |||  |:| |||
Db    281 ILYVFFDSPYSKSALCTYSMNTIKQSFSTSKLEGYTKQLPSPASGICLPAGKVVPHTTFE 340

Qy    378 VADSHPEVAQRVEPM 392
             ::: ||:||||||||
Db    341 VIEKYNVLDDIIKPL 355
```

```
Search completed: March 14, 2003, 09:26:00
Job time : 12.6906 secs
```